

it suitable for acquiring complete DNA genotyping of degraded sample [11]. The previous validation results of this new amplification system indicated that this system was accurate, specific, reproducible, robust, and the contained autosomal 57 InDels exhibited high polymorphisms in Chinese several populations [11,14–17]. Based on the Scientific Working Group on DNA Analysis Methods (SWG-DAM), one of the approaches to assess the availability of these loci for forensic DNA analysis is population genetic survey.

The Hani group is a traditional ethnic minority mainly distributed in China, Thailand, Laos, Vietnam, and Burma. Chinese Hani group is one of the formally recognized fifty-six ethnic groups. The result of the seventh national census in 2020 demonstrated that the Chinese Hani group had a population of 1.73 million (<http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>). Over 90% of Chinese Hani people live on the high plateau of the Yunnan province in Southern China. The language spoken by most of the Hani people is a member of the Yi language of the Tibeto-Burman branch and the Sino-Tibetan language family. Prior to 1949, the Hani people kept records by notch-carving sticks since they lacked their own script. Based on the Latin alphabet, the central government assisted them to develop the script in 1957 (http://www.yunnanadventure.com/minorities/Hani-Ethnic-Minority_11.html). In-depth systemic investigations of the genetic background and architecture of Chinese Yunnan Hani group were very important forensic significance. However, the genetic makeup of modern Yunnan Hani ethnic group still needs to be further explored, and few studies based on InDel polymorphic markers have been conducted in Chinese Yunnan Hani group [18]. In the current study, we utilized the novel 57 autosomal InDels to further validate the forensic property of this novel amplification system and explore the genetic polymorphisms of the 57 autosomal InDels in Chinese Yunnan Hani group. In addition, a variety of statistical analyses of the 57 autosomal InDels were used to probe the genetic differentiations and relationships between Chinese Yunnan Hani group and 30 worldwide reference populations.

2. Materials and Method

2.1 Sample Collections and Reference Populations

After obtaining written informed consents, blood samples of 221 unrelated healthy individuals were gathered from the Hani group residing in the Yunnan province of China. According to self-reported statements, all the participants included in present study had no common ancestors within three generations, and no migration events occurred in their family history. The sample collections strictly observed the principles of human and ethical research and were approved by the ethics committee of Xi'an Jiaotong University Health Science Center (Approval No. 2019-1039). In the 4008 individuals from 30 reference populations, 26 worldwide populations were from the 1000

Genomes Project Phase 3 database, and the remaining four populations were from previously published studies [14–17,19]. The detailed information and corresponding abbreviations for 31 populations are listed in **Supplementary Table 1**.

2.2 DNA Extraction, PCR Amplification and InDel Genotyping

The Chelex-100 method was used to extract genomic DNA. The multiplex PCR amplification of the novel panel was conducted on the GeneAmp PCR system 9700 Thermal Cycler (Thermo Fisher Scientific, Foster City, CA, USA) with the corresponding reaction system and amplification condition as previously reported [15]. Subsequent PCR amplification products were separated and detected using the ABI 3500xL Genetic Analyzer (Thermo Fisher Scientific, Foster City, CA, USA). GeneMapper ID-X software v1.5 (Thermo Fisher Scientific, Foster City, CA, USA) was used to implement the InDel genotyping. The positive and negative controls during the experimental procedures were DNA 9947A as well as 9948, and deionized water, respectively.

2.3 Statistical Analysis

The allele frequencies and forensic parameter values of autosomal 57 InDels containing the polymorphism information content (PIC), match probability (MP), power of discrimination (PD), probability of exclusion (PE), observed heterozygosity (Hobs), and expected heterozygosity (Hexp) in Chinese Yunnan Hani group and reference East Asian populations were measured by the STRAF online program (version v1.0.5, Gouy & Zieger, Bern, Switzerland) and visualized by raincloud plots by *R* software (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria) [20]. The values of cumulative power of discrimination (CPD) and probability of exclusion (CPE) of the 57 autosomal InDels in Chinese Yunnan Hani group were performed by the corresponding formula. The Hardy-Weinberg equilibrium (HWE) tests for the 57 autosomal InDels and linkage disequilibrium (LD) analyses for pairwise InDels in Chinese Yunnan Hani group were also conducted by the STRAF online program (version v1.0.5, Gouy & Zieger, Switzerland) [20]. The pairwise relationship tests involving full siblings and half siblings were simulated by Familias software (version 3, Kling & Tillmar & Egeland, Norwegian, Sweden) based on the allele frequencies of 57 autosomal InDels in Chinese Yunnan Hani group, and the results of likelihood ratio (LR) distributions were plotted with the density plot function via *R* software [21]. Locus by locus analysis of molecular variance (AMOVA) between Chinese Yunnan Hani group and 30 reference global populations together with pairwise fixation index (F_{ST}) values among the 31 worldwide populations were analyzed by the Arlequin software (version 3.5.1.2, Excoffier & Lischer, Lausanne, Switzerland) [22]. The D_A distances of pairwise populations were estimated by the DISPAN program (Nei

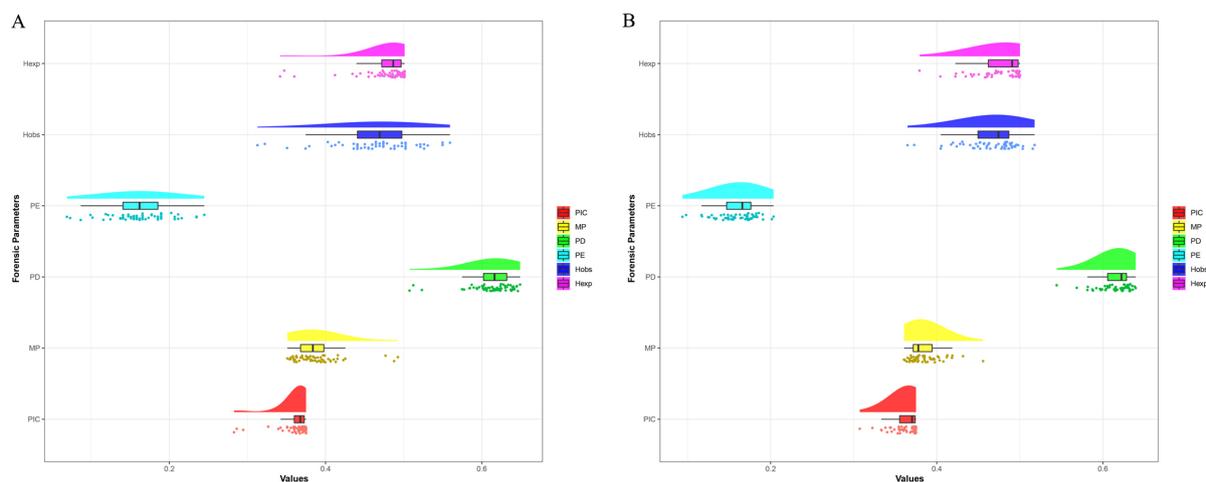


Fig. 1. The raincloud charts of forensic parameters of 57 autosomal InDels. (A) The raincloud chart of forensic parameters of 57 autosomal InDels in Chinese Yunnan Hani group. (B) The raincloud chart of forensic parameters of 57 autosomal InDels in East Asian populations including Chinese Yunnan Hani group.

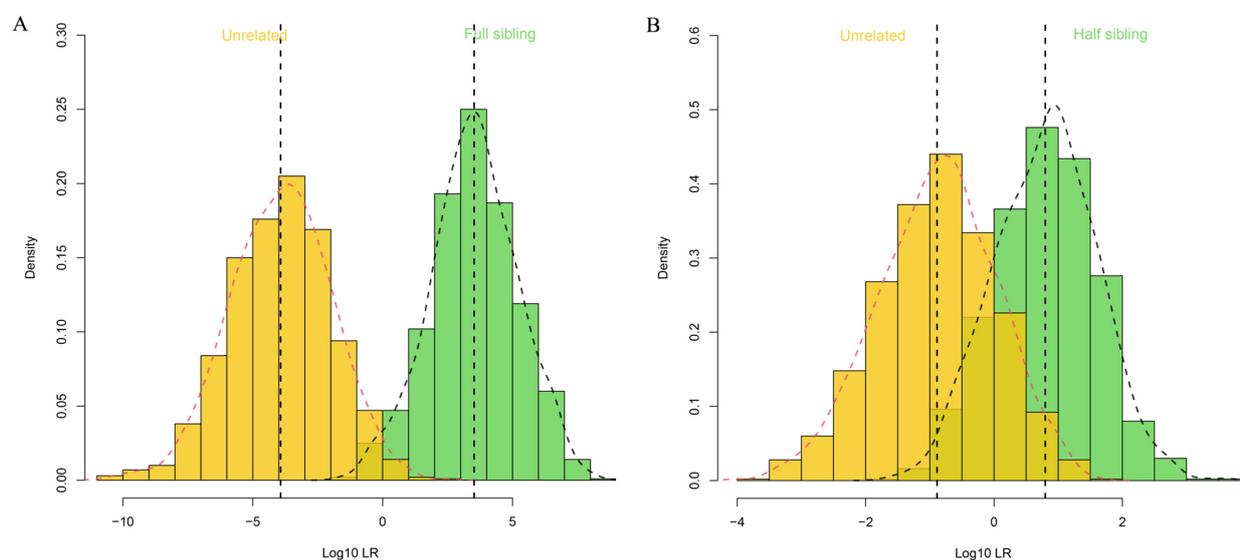


Fig. 2. The results of simulated full sibling and half sibling tests. (A) Log_{10}LR distribution plot to distinguish full siblings from unrelated individuals in Chinese Yunnan Hani group when using 57 autosomal InDels. (B) Log_{10}LR distribution plot to distinguish half siblings from unrelated individuals in Chinese Yunnan Hani group when using 57 autosomal InDels.

tion allele frequency distributions among the five reference intercontinental populations. For example, the rs3076465, rs71852971, rs3067397, rs72031009, and rs34419736 loci exhibited relatively large insertion allele frequencies in African populations but small insertion allele frequencies in other four intercontinental populations. The rs79225518, rs113011930, rs145010051, rs146875868, and rs77635204 loci embodied relatively high insertion allele frequencies in East Asian populations but low insertion allele frequencies in other four intercontinental populations.

Following Bonferroni correction, the significance threshold was changed to 0.0012 ($p = 0.05/57 = 0.0008771$). The locus-by-locus p values computed using the AMOVA approach are listed in **Supplementary Table 6**. The Yun-

nan Hani group showed the minimum amount of significant differences at 2–30 loci with East Asian populations, but the maximum number of significant differences at 32–41 loci with African populations. In addition, the significant differences among Chinese Yunnan Hani group and the remaining populations were observed to be 16–30 loci in four American populations, 36–41 loci with five European populations, and 25–27 loci with five South Asian populations.

The pairwise F_{ST} values and D_A distances among Chinese Yunnan Hani group and 30 reference populations are displayed in **Supplementary Tables 7,8**, respectively. Compared with other populations, Chinese Yunnan Hani group showed the minimum F_{ST} value with KHV (0.0090), followed by CHS (0.0095) and CDX (0.0103). The max-



Fig. 3. The heatmap of insertion allele frequencies of 57 autosomal InDels in the Chinese Yunnan Hani group and 30 reference populations. ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Tuscans in Italy; BEB, Bengali in Bangladesh; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; PJL, Punjabi in Lahore, Pakistan; STU, Sri Lankan Tamil in the UK; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese, China; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; SHP, Dingjie Sherpa, China; HNL, Hainan Li, China; HNH, Hunan Han, China; YNM, Yunnan Miao, China; YNH, Yunnan Hani, China.

imum F_{ST} value was observed between Chinese Yunnan Hani group and YRI (0.2267), followed by the GWD (0.2236) and ESN (0.2206). The D_A distances among Chinese Yunnan Hani group and 30 reference populations exhibited the similar results with the F_{ST} values. The closest D_A distance was observed between Chinese Yunnan Hani group and KHV (0.0032), followed by the HNL (0.0032) and CHS (0.0033); while the farthest D_A distance was between Chinese Yunnan Hani group and YRI (0.1014), followed by ESN (0.0999) and MSL (0.0967). Two heatmaps constructed based on the values of pairwise F_{ST} and D_A distances are exhibited through different colors in Fig. 4A,C, respectively. As shown in the heatmaps, Chinese Yunnan Hani group had larger F_{ST} values and D_A distances with the African populations but smaller F_{ST} values and D_A distances with reference East Asian populations. The pairwise F_{ST} values and D_A distances between Chinese Yunnan Hani group and 30 reference populations were also displayed in the line graphs in Fig. 4B,D, respectively.

The pairwise F_{ST} values of five intercontinental populations in pairs, representing the same intercontinental pop-

ulations as a whole, were also calculated, and the results are presented in **Supplementary Table 9** and visualized through a boxplot in Fig. 5, respectively. The rs3076465, rs71852971, rs59841142, rs3067397, rs72031009, and rs34419736 loci exhibited relatively high F_{ST} values ($F_{ST} > 0.15$) between African populations and other intercontinental populations. The rs145577149, rs113011930, rs145010051, rs76158822, and rs77635204 loci showed relatively high F_{ST} values ($F_{ST} > 0.15$) between East Asian populations and other intercontinental populations, implying that these loci might be informative for discriminating East Asian populations from other intercontinental populations.

3.3 Phylogenetic Relationship Reconstruction between Chinese Yunnan Hani Group and 30 Reference Populations

Two phylogenetic trees, one rooted tree in Fig. 6A and one unrooted tree in Fig. 6B, were also used to assess the genetic relationships of the 31 populations. As depicted in Fig. 6A, the loop phylogenetic tree created using the pair-

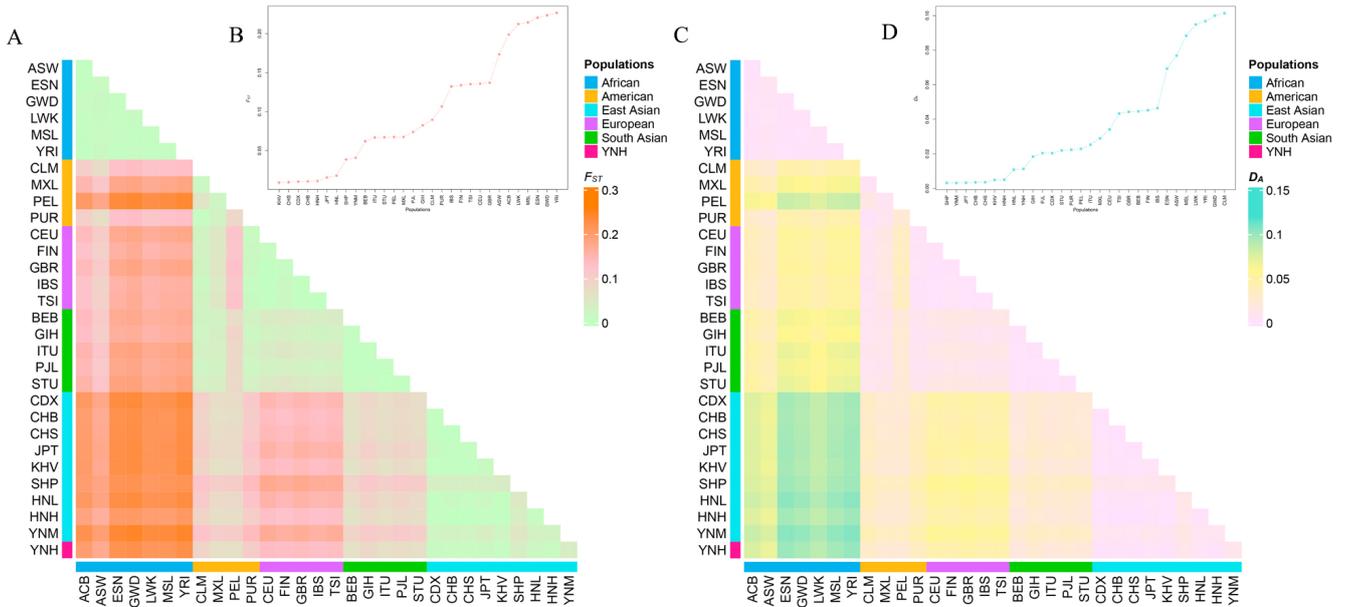


Fig. 4. The heatmaps and the line graphs of the pairwise F_{ST} values and D_A distances. (A) The heatmap of pairwise F_{ST} values among Chinese Yunnan Hani group and 30 reference populations. (B) The line graph of the pairwise F_{ST} values between Chinese Yunnan Hani group and 30 reference populations. (C) The heatmap of the pairwise D_A distances among Chinese Yunnan Hani group and 30 reference populations. (D) The line graph of the pairwise D_A distances between Chinese Yunnan Hani group and 30 reference populations.

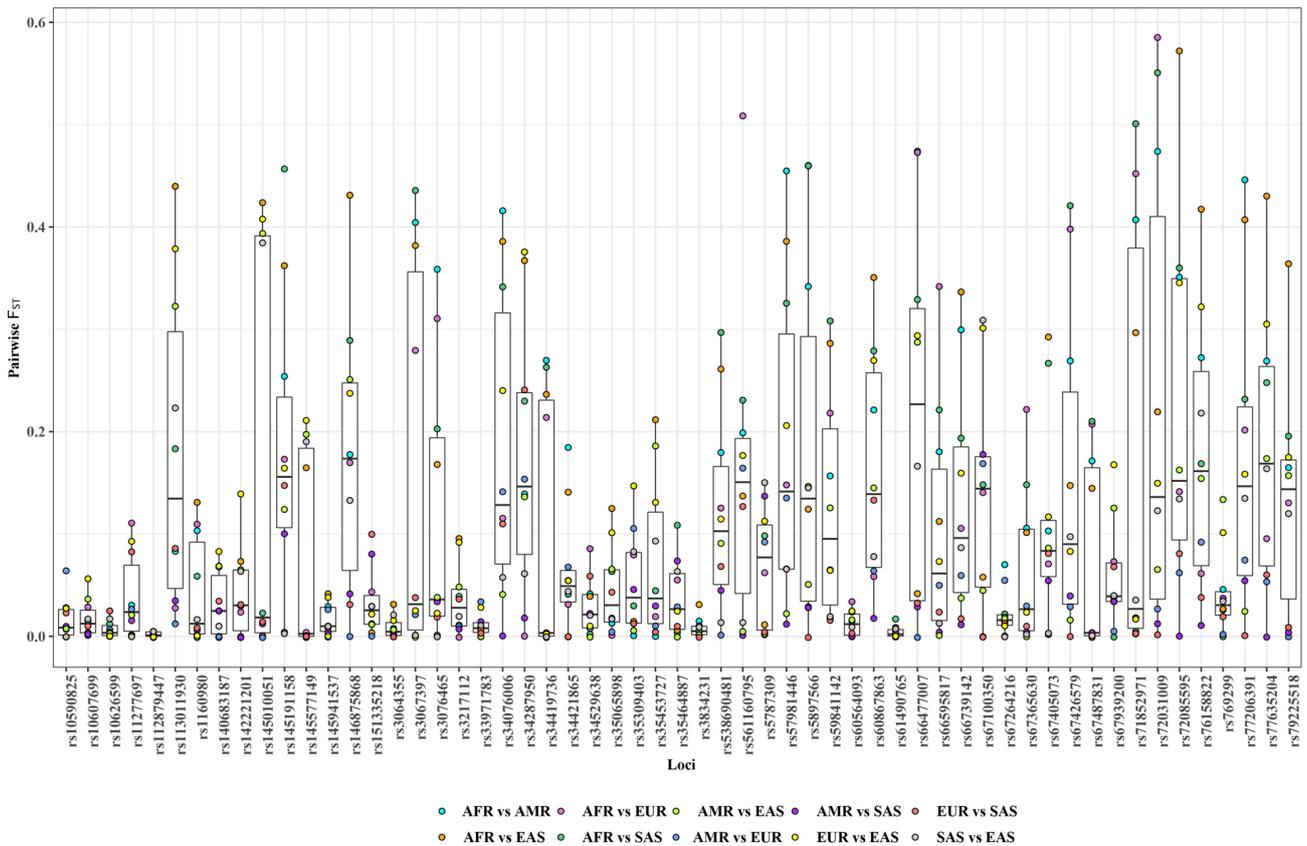


Fig. 5. The pairwise F_{ST} values at 57 autosomal InDels among pairwise intercontinental populations.

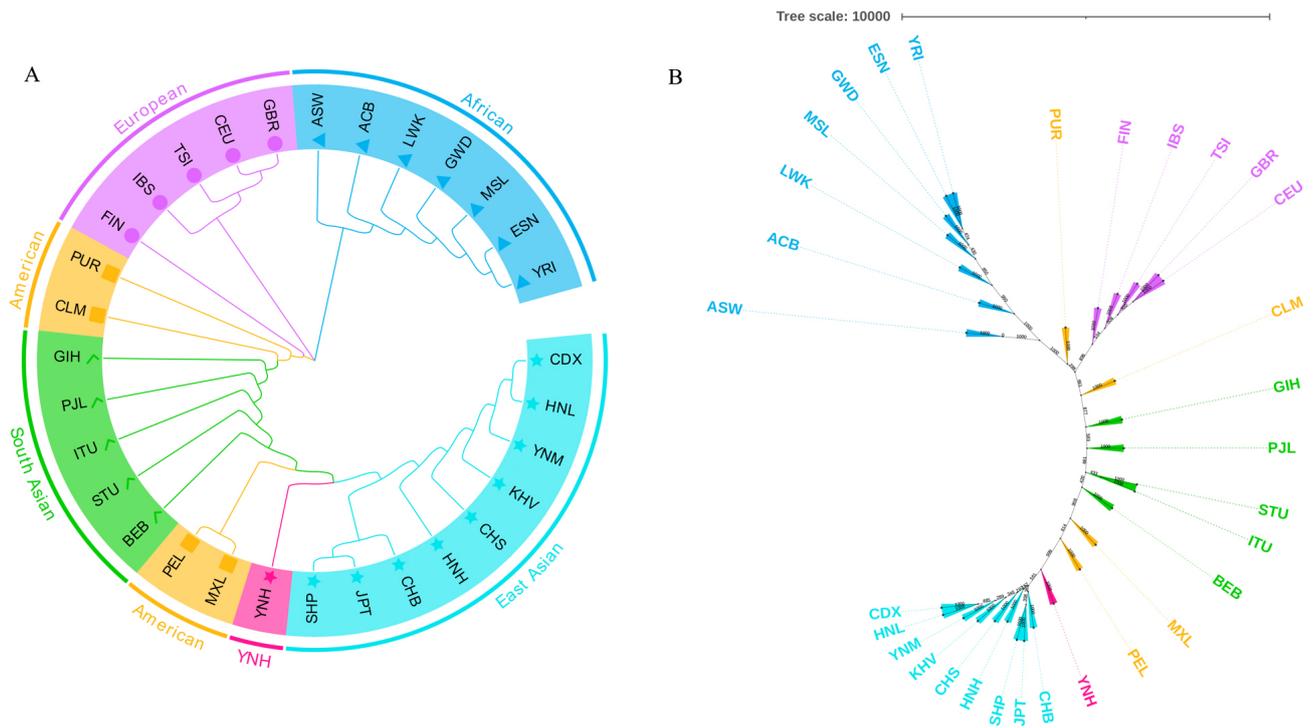


Fig. 6. The phylogenetic relationship reconstructions among Chinese Yunnan Hani group and 30 reference populations. (A) The phylogenetic tree conducted based on the pairwise D_A distances among Chinese Yunnan Hani group and 30 reference populations. (B) The unrooted tree developed based on the allele frequencies of 57 autosomal InDels among Chinese Yunnan Hani group and 30 reference populations.

wise D_A distances was mostly separated into two branches: seven populations from Africa were classified into the first branch, and Chinese Yunnan Hani group and other populations made up the second branch. In the second branch, four populations from Europe formed into one subbranch while the rest populations grouped into another subbranch. Chinese Yunnan Hani group clustered with nine reference East Asian populations. These trees indicated that the genetic relationships among Chinese Yunnan Hani group and reference East Asian populations were closer than those of the other intercontinental populations. A similar branch distribution pattern could be observed in the unrooted tree in Fig. 6B, which was derived from the allelic frequencies of the 57 autosomal InDels in the 31 worldwide populations.

3.4 PCA and MDS Analyses among Chinese Yunnan Hani Group and 30 Reference Populations

The PCA and MDS analyses were also employed to further investigate the genetic relationships among Chinese Yunnan Hani group and 30 worldwide reference populations, which are exhibited in Fig. 7. The results of genotype-based PCA plots at the individual level are exhibited in Fig. 7A–D, while the results of allele frequency-based PCA chart at the population level is presented in Fig. 7E. The African, reference East Asian, and European populations was in Fig. 7A, while South Asian populations, American populations, and Chinese Yunnan Hani

group are shown in Fig. 7B–D, respectively. The individuals from African, European and reference East Asian populations could be clearly distinguished. Most of Chinese Yunnan Hani individuals overlapped with reference East Asian populations. At the population level in Fig. 7E, the first two principal components contributed cumulatively 69.50% of the total variation, with PC1 and PC2 accounting for 46.29%, and 23.21%, respectively. PC1 could categorize African, reference East Asian populations, and Chinese Yunnan Hani group from the rest of the populations; PC2 mainly separated African, European populations from other populations. The \cos^2 values of 57 autosomal InDels are shown in Fig. 8, and the rs112879447, rs10626599, rs61490765, rs151335218, and rs10590825 loci are located in the relatively inner part of the ring and presented relatively low \cos^2 values. The result of the MDS analysis in terms of pairwise F_{ST} values is shown in Fig. 7F. A similar pattern of population distributions was observed in the MDS plot, which further confirmed the close relationships between Chinese Yunnan Hani group and reference East Asian populations involved in this study.

3.5 Population Genetic Structure Analyses between Chinese Yunnan Hani Group and 30 Reference Populations

To further demonstrate the ancestral compositions of Chinese Yunnan Hani group, population genetic structure

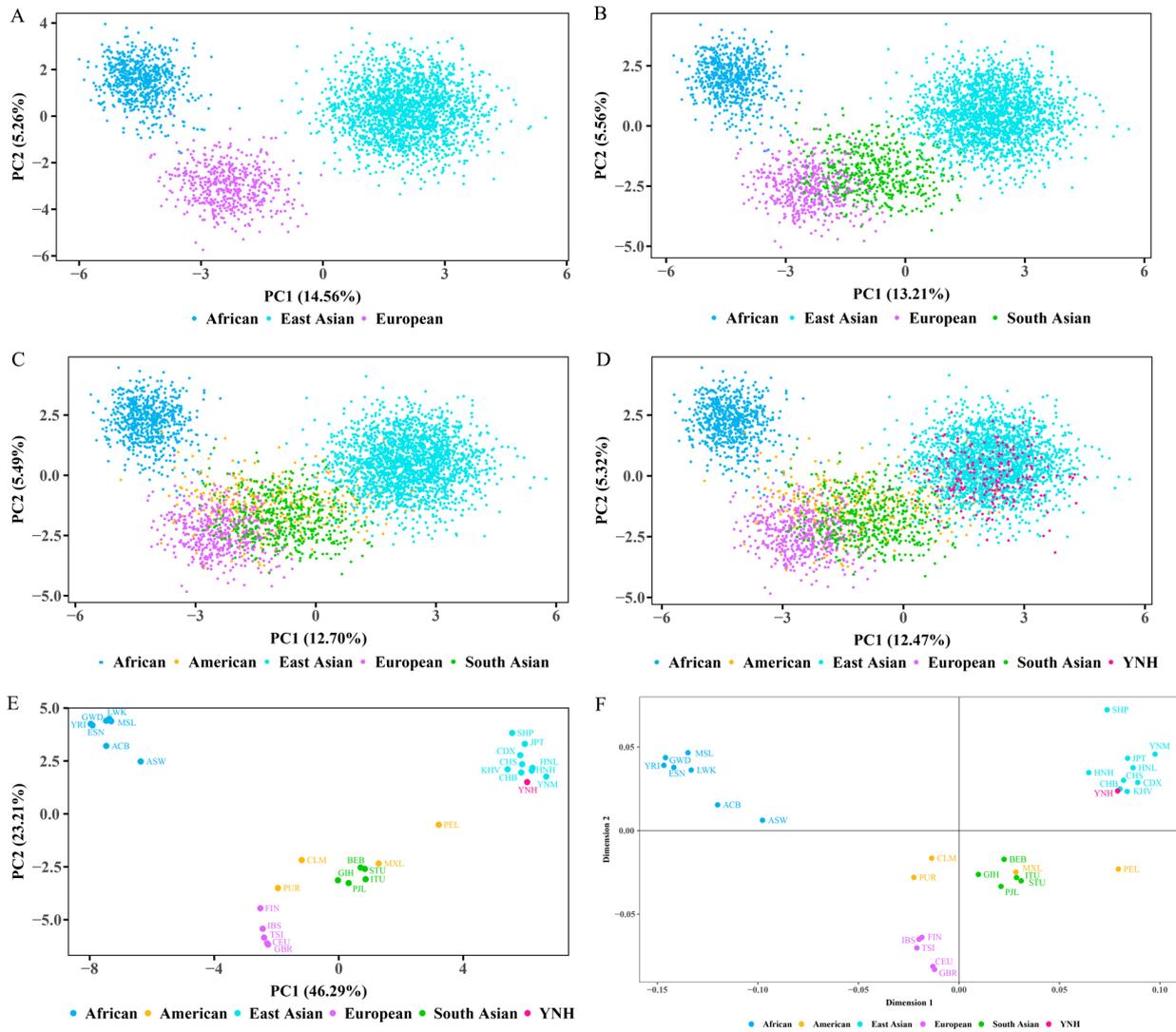


Fig. 7. The results of PCA and MDS analyses among Chinese Yunnan Hani group and reference populations. (A) The PCA plot on the individual level in 21 reference populations from three continents. (B) The PCA plot on the individual level in 26 reference populations from four continents. (C) The PCA plot on the individual level in 30 reference populations from five continents. (D) The PCA plot on the individual level among Chinese Yunnan Hani group and 30 reference populations. (E) The PCA plot on the population level among Chinese Yunnan Hani group and 30 reference populations. (F) The MDS analysis plot based on the pairwise F_{ST} values among Chinese Yunnan Hani group and 30 reference populations.

analyses were also performed, and the results are presented in Fig. 9 and **Supplementary Figs. 1,2**, respectively. As shown in **Supplementary Fig. 1**, the Delta K maximize was observed at $K = 3$, which meant the optimum K value estimated by the Structure Harvester Website was three. **Supplementary Fig. 2** showed the locations of the 31 global reference populations on the world map and the structure analysis results ($K = 3$) at the 31 population level through the pie plots. The estimated ancestral proportions of Chinese Yunnan Hani group and 30 reference populations by assuming three ancestral populations are shown in the boxplot in Fig. 9A. The structure analysis results ($K = 3$) at 4229 individual levels are exhibited in the circle plot in Fig. 9B and the triangle plot in Fig. 9C, respectively. The

dots representing Chinese Yunnan Hani individuals overlapped for the most part with those representing the reference East Asian populations in Fig. 9C. Moreover, the detailed ancestral proportion ($K = 3$) of Chinese Yunnan Hani group and other 30 reference populations were also presented through the circle chart in Fig. 9D. When $K = 3$, three ancestry compositions of the African, East Asian, and European clusters were further divided, and the dominant ancestry component in Chinese Yunnan Hani group was East Asian. More specifically, the percentages of African, European, and East Asian ancestry components in Chinese Yunnan Hani group were 0.0146, 0.0455, and 0.9399, respectively. Ancestral compositions of Chinese Yunnan Hani group were always in accord with other East Asian popula-

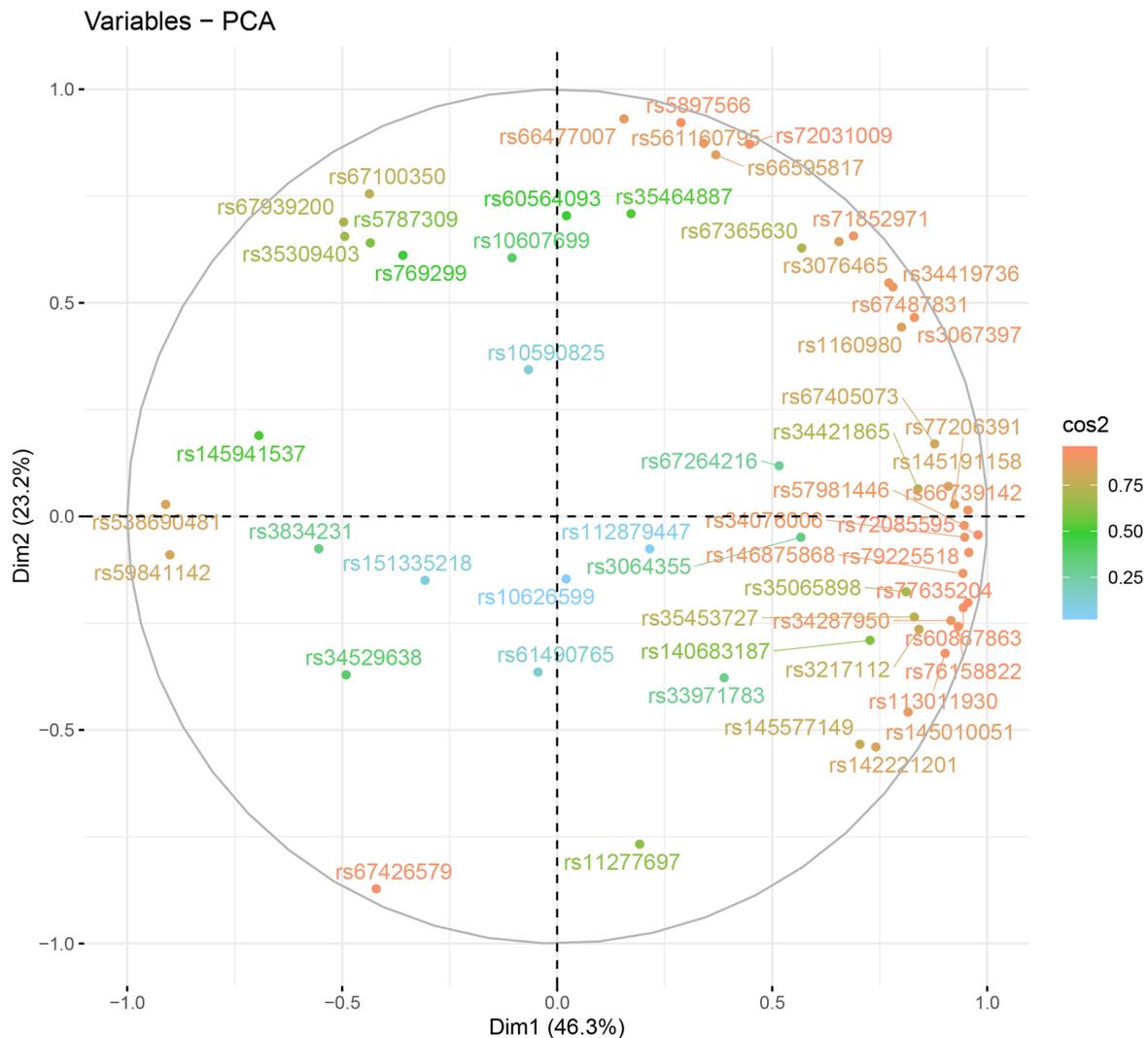


Fig. 8. The cos2 values on the basis of allelic frequencies of 57 autosomal InDels in Chinese Yunnan Hani group and 30 reference populations.

tions, indicating that the genetic architecture between Chinese Yunnan Hani group and East Asian populations was more similar than those of other intercontinental populations.

4. Discussion

InDels, a new generation of genetic markers, have unique advantages in forensic genetic analysis and population genetics, since they combine the advantages of STRs and SNPs [31]. However, most InDels have some limitations as diallelic genetic markers. First, the genetic information carried by InDels is limited, and more InDel loci need to be combined to achieve the high discrimination efficacy. The combination of a large number of loci in the same multiplex amplification system increases the technical difficulties. Second, the allelic frequencies of InDel loci show the distribution differences in different populations, so more population genetic data are needed to investigate

and support the actual application in the practice of forensic DNA identification [32]. Therefore, it is crucial to construct a multiple amplification system containing more and higher genetic polymorphism InDel loci, and also to test their forensic efficacies in different populations.

The formation of the Hani ethnic group has taken place over a long historical period and integrated various cultural factors, making the issue of its ethnic background very complex. The origin of the Hani is not precisely known, but many academics believed that the Hani group, Yi and Lahu ethnic minorities all have common ancestors. And they all descended from the ancient Qiang people, a nomadic clan that once inhabited the Qinghai-Tibetan Plateau and gave rise to numerous ethnic groups across southern and southwestern China (http://www.yunnanadventure.com/minorities/Hani-Ethnic-Minority_11.html). Due to population migrations, the Hani ethnic group comprises many subgroups. Studying Chinese Hani group not only enriches the genetic data of the Hani group and further explores the genetic

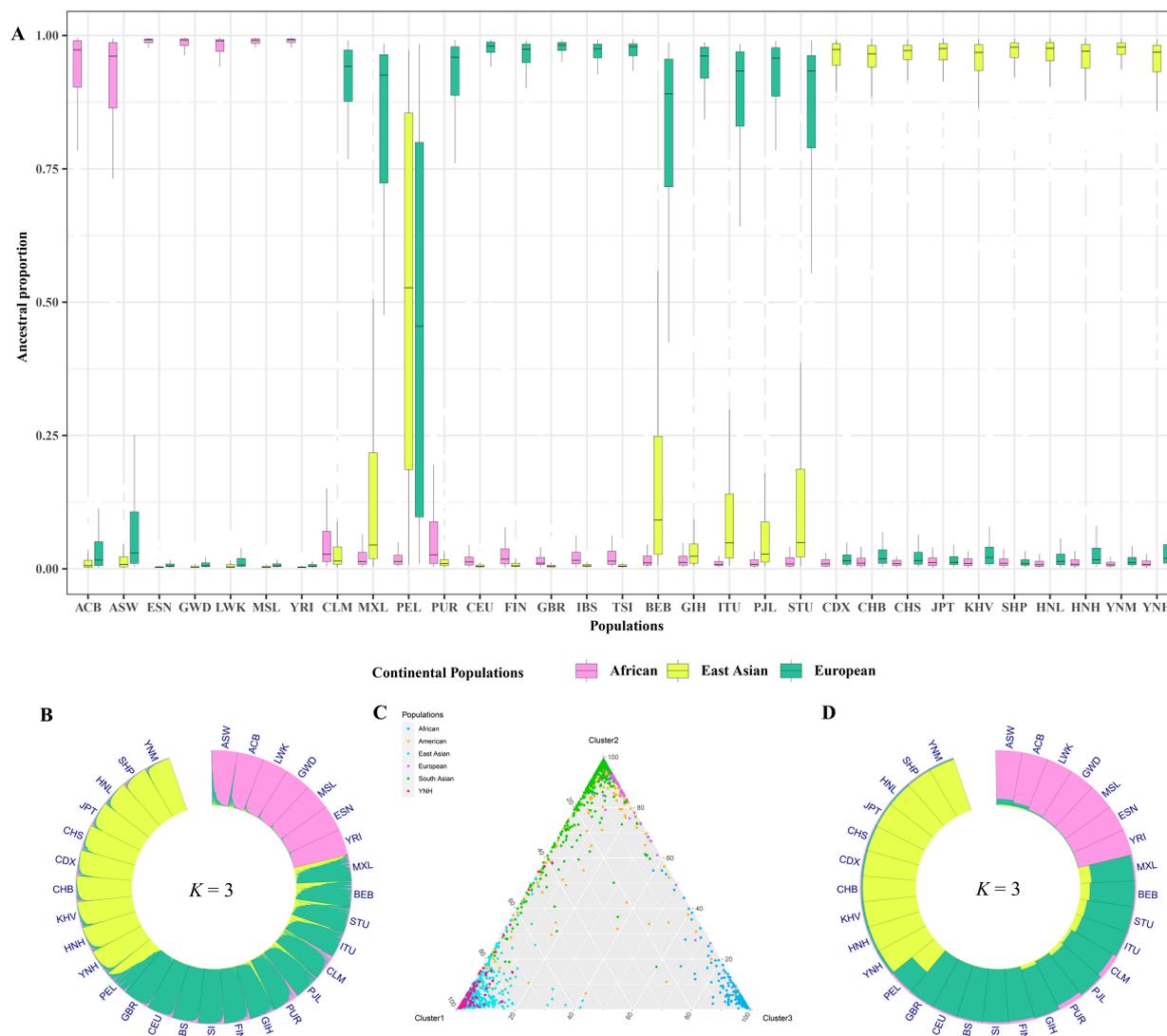


Fig. 9. The outcomes of population genetic structure analyses between Chinese Yunnan Hani group and 30 reference populations.

(A) Estimated ancestry proportions of Chinese Yunnan Hani group and 30 reference populations by assuming three ancestral populations. (B) The population genetic structure analysis results ($K = 3$) on the individual level among Chinese Yunnan Hani group and 30 reference populations. (C) Clustering analysis results ($K = 3$) for individual ancestry estimation among Chinese Yunnan Hani group and 30 reference populations. (D) The population genetic structure analysis results ($K = 3$) on the population level among Chinese Yunnan Hani group and 30 reference populations.

structure of the Hani group, but also lays the foundation for further forensic applications of InDels.

In this study, we assessed the genetic polymorphisms and forensic application efficiencies of the novel panel containing 57 autosomal InDels in Chinese Yunnan Hani group. No significant HWE deviation was observed at the 57 autosomal InDels, and no linkage disequilibrium was detected in the pairs of 57 autosomal InDels, which indicated that these samples employed in this study were representative in Chinese Yunnan Hani group and offered the foundation for the computations of forensic parameters and analyses of population genetic relationships. In a population, observed heterozygosity refers to the percentage of heterozygotes in

all individual genotypes of a genetic marker. The high degree of heterozygosity indicates that the genetic marker is high application value in forensic personal identification. PIC is applied to estimate the extent of polymorphism in a genetic marker. When the PIC value exceeds 0.25, it indicates that valid genetic information can be provided [33,34]. PD is a metric used to evaluate the degree of effectiveness of a genetic marker in identifying different individuals. PE is an index used to appraise the practical value of a particular genetic marker in parentage identification case. In Chinese Yunnan Hani group, all the 57 autosomal InDels showed heterozygosity values above 0.3, PIC values higher than 0.25, and mean PD and PE values of 0.6118 and

Ethics Approval and Consent to Participate

The present research strictly adhered to the ethical guidelines of the Helsinki Declaration and was approved by the ethics committee of Xi'an Jiaotong University, China (Approval No. 2019-1039).

Acknowledgment

The authors want to thank the volunteers in this research.

Funding

This research was funded by the National Natural Science Foundation of China, grant number 81930055.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2801005>.

References

- [1] Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research*. 2006; 16: 1182–1190.
- [2] Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*. 2009; 30: 3682–3690.
- [3] LaRue BL, Ge J, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. *International Journal of Legal Medicine*. 2012; 126: 533–540.
- [4] Jin X, Shen C, Chen C, Guo Y, Cui W, Wang Y, *et al.* Ancestry informative DIP loci for dissecting genetic structure and ancestry proportions of Qinghai Tibetan and Tibet Tibetan groups. *Molecular Biology Reports*. 2020; 47: 1079–1087.
- [5] Xie T, Shen C, Jin X, Lan Q, Fang Y, Zhu B. Genetic Structural Differentiation Analyses of Intercontinental Populations and Ancestry Inference of the Chinese Hui Group Based on a Novel Developed Autosomal AIM-InDel Genotyping System. *BioMed Research International*. 2020; 2020: 2124370.
- [6] Oldoni F, Castella V, Hall D. Application of DIP-STRs to sexual/physical assault investigations: Eight case reports. *Forensic Science International: Genetics*. 2017; 30: 106–113.
- [7] Wilkins S, Zhang KW, Mahfuz I, Quantin R, D’Cruz N, Hutson J, *et al.* Insertion/deletion polymorphisms in the Δ Np63 promoter are a risk factor for bladder exstrophy epispadias complex. *PLoS Genetics*. 2012; 8: e1003070.
- [8] Chen L, Du W, Wu W, Yu A, Pan X, Feng P, *et al.* Developmental validation of a novel six-dye typing system with 47 a-InDels and 2 Y-InDels. *Forensic Science International: Genetics*. 2019; 40: 64–73.
- [9] Cui W, Jin X, Guo Y, Chen C, Zhang W, Kong T, *et al.* An innovative panel containing a set of insertion/deletion loci for individual identification and its forensic efficiency evaluations in Chinese Hui ethnic minority. *Molecular Genetics and Genomic Medicine*. 2020; 8: e1074.
- [10] Zhang X, Shen C, Jin X, Guo Y, Xie T, Zhu B. Developmental validations of a self-developed 39 AIM-InDel panel and its forensic efficiency evaluations in the Shaanxi Han population. *International Journal of Legal Medicine*. 2021; 135: 1359–1367.
- [11] Liu J, Du W, Jiang L, Liu C, Chen L, Zheng Y, *et al.* Development and validation of a forensic multiplex InDel assay: the AGCU InDel 60 kit. *Electrophoresis*. 2022; 43: 1871–1881.
- [12] Alaeddini R, Walsh SJ, Abbas A. Forensic implications of genetic analyses from degraded DNA—a review. *Forensic Science International: Genetics*. 2010; 4: 148–157.
- [13] Takahashi M, Kato Y, Mukoyama H, Kanaya H, Kamiyama S. Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues—correlation between the size of the alleles and that of the template DNA. *Forensic Science International*. 1997; 90: 1–9.
- [14] Chen X, Nie S, Hu L, Fang Y, Cui W, Xu H, *et al.* Forensic efficacy evaluation and genetic structure exploration of the Yunnan Miao group by a multiplex InDel panel. *Electrophoresis*. 2022; 43: 1765–1773.
- [15] Fang Y, Zhao C, Jin X, Lan Q, Lan J, Xie T, *et al.* Genetic characterization evaluation of a novel multiple system containing 57 deletion/insertion polymorphic loci with short amplicons in Hunan Han population and its intercontinental populations analyses. *Gene*. 2022; 809: 146006.
- [16] Wang M, Du W, Tang R, Liu Y, Zou X, Yuan D, *et al.* Genomic history and forensic characteristics of Sherpa highlanders on the Tibetan Plateau inferred from high-resolution InDel panel and genome-wide SNPs. *Forensic Science International: Genetics*. 2022; 56: 102633.
- [17] Fan H, He Y, Li S, Xie Q, Wang F, Du Z, *et al.* Systematic Evaluation of a Novel 6-dye Direct and Multiplex PCR-CE-Based InDel Typing System for Forensic Purposes. *Frontiers in Genetics*. 2022; 12: 744645.
- [18] Cui W, Nie S, Fang Y, Chen M, Zhao M, Lan Q, *et al.* Insights into AIM-InDel diversities in Yunnan Miao and Hani ethnic groups of China for forensic and population genetic purposes. *Hereditas*. 2022; 159: 22.
- [19] Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, *et al.* A global reference for human genetic variation. *Nature*. 2015; 526: 68–74.
- [20] Gouy A, Zieger M. STRAF—a convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*. 2017; 30: 148–151.
- [21] Kling D, Tillmar AO, Egeland T. Familias 3 – Extensions and new functionality. *Forensic Science International: Genetics*. 2014; 13: 121–127.
- [22] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010; 10: 564–567.
- [23] Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*. 1983; 19: 153–170.
- [24] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016; 33: 1870–1874.
- [25] Shimada MK, Nishida T. A modification of the PHYLIP program: a solution for the redundant cluster problem, and an implementation of an automatic bootstrapping on trees inferred from original data. *Molecular Phylogenetics and Evolution*. 2017; 109: 409–414.
- [26] Rousset F. Genepop’007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*. 2008; 8: 103–106.
- [27] Evanno G, Regnaut S, Goudet J. Detecting the number of clus-

- ters of individuals using the software structure: a simulation study. *Molecular Ecology*. 2005; 14: 2611–2620.
- [28] Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 2012; 4: 359–361.
- [29] Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007; 23: 1801–1806.
- [30] Feng Q, Lu D, Xu S. AncestryPainter: a Graphic Program for Displaying Ancestry Composition of Populations and Individuals. *Genomics, Proteomics and Bioinformatics*. 2018; 16: 382–385.
- [31] Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, Carracedo A, *et al.* Forensic performance of two insertion–deletion marker assays. *International Journal of Legal Medicine*. 2012; 126: 725–737.
- [32] Sheng X, Bao Y, Zhang JS, Li M, Li YN, Xu QN, *et al.* Research Progress on InDel Genetic Marker in Forensic Science. *Fa Yi Xue Za Zhi*. 2018; 34: 420–427.
- [33] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978; 89: 583–590.
- [34] Shete S, Tiwari H, Elston RC. On Estimating the Heterozygosity and Polymorphism Information Content Value. *Theoretical Population Biology*. 2000; 57: 265–271.
- [35] Zhang X, Du L, Huang Z, Gu T, Hu L, Nie S. Genetic variation of 20 autosomal STR loci in three ethnic groups (Zhuang, Dai and Hani) in the Yunnan province of southwestern China. *Forensic Science International: Genetics*. 2017; 31: e41–e42.
- [36] Enoch M, Shen P, Xu K, Hodgkinson C, Goldman D. Using ancestry-informative markers to define populations and detect population stratification. *Journal of Psychopharmacology*. 2006; 20: 19–26.
- [37] Jin X, Cui W, Chen C, Guo Y, Tao Y, Lan Q, *et al.* Biogeographic origin prediction of three continental populations through 42 ancestry informative SNPs. *Electrophoresis*. 2020; 41: 235–245.
- [38] Wang HY, Hu YH, Cao YY, Zhu Q, Huang YG, Li X, *et al.* AI-SNPs screening based on the whole genome data and research on genetic structure differences of subcontinent populations. *Yi Chuan*. 2021; 43: 938–948.
- [39] Kersbergen P, van Duijn K, Kloosterman AD, den Dunnen JT, Kayser M, de Knijff P. Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics*. 2009; 10: 69.
- [40] Jakobsson M, Edge MD, Rosenberg NA. The relationship between F(ST) and the frequency of the most frequent allele. *Genetics*. 2013; 193: 515–528.
- [41] Huang Y, Yao J, Li J, Wen J, Yuan X, Xu B. Population genetic data for 17 autosomal STR markers in the Hani population from China. *International Journal of Legal Medicine*. 2015; 129: 995–996.
- [42] Hu L, Gu T, Fan X, Yuan X, Rao M, Pang JB, *et al.* Genetic polymorphisms of 24 Y-STR loci in Hani ethnic minority from Yunnan Province, Southwest China. *International Journal of Legal Medicine*. 2017; 131: 1235–1237.
- [43] Liu L, Yao J, Huang Y, Gao L, Dai J, Yuan X, *et al.* Genetic polymorphisms of 16 X-STR loci in the Hani population from Southwest China. *Forensic Sciences Research*. 2022; 7: 196–201.